

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-187559

(43)Date of publication of application : 04.07.2000

(51)Int.Cl.

G06F 3/06

(21)Application number : 10-366231

(71)Applicant : HITACHI LTD

(22)Date of filing : 24.12.1998

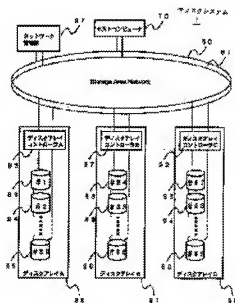
(72)Inventor : MATSUMOTO JUN  
MATSUNAMI NAOTO  
YAGISAWA IKUYA  
YAMAMOTO MASAYUKI  
TAKAMOTO KENICHI

## (54) DISK SYSTEM

## (57)Abstract:

**PROBLEM TO BE SOLVED:** To provide a disk system which avoids the drop of throughput due to controller load increase caused by the increase of the number of disks.

**SOLUTION:** Plural disk arrays 86, 91 and 96 are connected to fiber channel loops 80 and 81, each disk array is subjected to function sharing as to whether a sequential command is preferentially processed or whether a random command is preferentially processed, a network managing part 97 decides whether a command from a host computer 70 is a sequential command or a random command, and the command is made to be processed by a disk array corresponding to the decision. Thus, it is possible to configure a disk drive of disk arrays in accordance with command characteristics and to realize high throughput performance at a low cost.



## LEGAL STATUS

[Date of request for examination]

29.05.2002

[Date of sending the examiner's decision of rejection]

11.01.2005

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(51) Int. Cl. <sup>7</sup>	識別記号	F I	ページ* (参考)
G 0 6 F 3/06	3 0 1	G 0 6 F 3/06	3 0 1 G 5 B 0 6 5
	3 0 2		3 0 2 E

審査請求 未請求 請求項の数5 O L (全 18 頁)

(21) 出願番号 特願平10-366231

(22) 出願日 平成10年12月24日(1998.12.24)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 松本 純

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所システム開発研究所内

(72) 発明者 松並 直人

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所システム開発研究所内

(74) 代理人 100095511

弁理士 有近 紳志郎

最終頁に続く

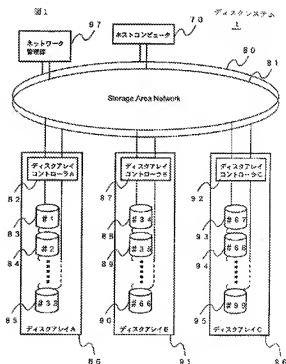
## (54) 【発明の名称】 ディスクシステム

## (57) 【要約】

【課題】 ディスク台数増加に伴うコントローラ負荷増大によるスループット性能の低下を回避するディスクシステムを提供する。

【解決手段】 ファイバチャネルのループ 80、81 に複数台のディスクアレイ 86、91、96 を接続し、各ディスクアレイに対してシーケンシャルコマンドを優先的に処理するかランダムコマンドを優先的に処理するかを機能分担させ、ホストコンピュータ 70 からのコマンドがシーケンシャルコマンドかランダムコマンドかをネットワーク管理部 97 で区分し、その区分に応じたディスクアレイに該コマンドを処理させる。

【効果】 コマンドの特性に応じてディスクアレイのディスク装置を構成する可能となり、低コストで高スループット性能を実現できる。



## 【特許請求の範囲】

【請求項1】 1台以上のホストコンピュータと、2台以上のディスクアレイとをファイバチャネルのループで接続したディスクシステムであって、

シーケンシャルコマンドを優先的に処理するディスクアレイかランダムコマンドを優先的に処理するディスクアレイかを識別しうる識別IDを前記ディスクアレイの少なくとも2台に対して決定する識別ID決定手段と、前記ホストコンピュータから前記ディスクアレイに対して送附したコマンドがシーケンシャルコマンドかランダムコマンドかを区分するコマンド区分手段と、シーケンシャルコマンドにはシーケンシャルコマンドを優先的に処理するディスクアレイを割り当てると共にランダムコマンドにはランダムコマンドを優先的に処理するディスクアレイを割り当てるためにコマンドに対して前記識別IDを付加する識別ID付加手段とを具備したことを特徴とするディスクシステム。

【請求項2】 1台以上のホストコンピュータと、1台以上のディスクアレイとをファイバチャネルのループで接続したディスクシステムであって、前記ディスクアレイの少なくとも1台は、2台以上のディスクアレイコントローラおよび2台以上のディスク装置を含み、

前記ディスクアレイコントローラおよび前記ディスク装置をグループ化するディスクグループ構成手段と、シーケンシャルコマンドを優先的に処理するグループかランダムコマンドを優先的に処理するグループかを識別しうる識別IDを前記グループに対して決定する識別ID決定手段と、前記ホストコンピュータから前記ディスクアレイに対して送附したコマンドがシーケンシャルコマンドかランダムコマンドかを区分するコマンド区分手段と、シーケンシャルコマンドにはシーケンシャルコマンドを優先的に処理するグループを割り当てると共にランダムコマンドにはランダムコマンドを優先的に処理するグループを割り当てるためにコマンドに対して前記識別IDを付加する識別ID付加手段とを具備したことを特徴とするディスクシステム。

【請求項3】 請求項1または請求項2に記載のディスクシステムにおいて、前記コマンド区分手段および前記識別ID付加手段を複数有し、前記コマンド区分処理および前記識別ID付加処理を多重実行可能としたことを特徴とするディスクシステム。

【請求項4】 1台以上のホストコンピュータと1台以上の高速アクセスのディスクアレイと1台以上の高信頼性のディスクアレイおよびDVDやMOなどの低速アクセス、大容量のライブラリライの少なくとも一方とをファイバチャネルのループで接続したディスクシステムであって、

アクセス頻度に従ってデータを前記高速アクセスのディスクアレイから前記高信頼性のディスクアレイまたは前記

低速アクセス・大容量のライブラリに移行するデータ移行手段と、移行後のデータのアドレスを管理するデータベース手段とを具備したことを特徴とするディスクシステム。

【請求項5】 請求項1から請求項4のいずれかに記載のディスクシステムにおいて、複数本のファイバチャネルのループの利用率を検出し利用率的低いループを優先的に用いてコマンドを送信するループ選択手段をさらに具備したことを特徴とするディスクシステム。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、ディスクシステムに関し、さらに詳しくは、ファイバチャネル・インタフェースを有するハードディスク、光ディスク、DVD、半導体ディスクなどのディスク装置及びそれらで構成されたディスクアレイをホストコンピュータと相互にネットワークで接続したディスクシステムに関する。

## 【0002】

【従来の技術】従来、コンピュータと磁気ディスク装置等の記憶媒体を接続するインタフェースとして SCSI (Small Computer System Interface) に替わり、より大容量データを高速に転送可能とするインタフェースとして、ファイバチャネルが注目されている。ファイバチャネルは、ホストコンピュータと周辺機器との間で大容量のデータを入出力するインタフェースであるチャネルというプロトコルと、ホストコンピュータと周辺機器がいつでも通信できる環境を提供するネットワークというプロトコルの2つの基本的なプロトコルから成り立っている。

【0003】ファイバチャネルを使用した機器をノードと呼ぶ。ノードは、他のノードのポートにアクセスするために1つ以上のポートを持っている。ANSI (American National Standards Institute) で規格化されている F C - A L (Fibre Channel Arbitrated Loop) では、各ノードが2つのポートを持ち、それぞれのポートが独立のループを構成する。

【0004】図18に、2つのループで各ノード間を F C - A L で相互接続したディスクアレイのシステム概念図を示す。ホスト70とディスクアレイコントローラ73とがループ71、72に接続されている。また、ディスクアレイコントローラ73の配下に、単体のディスク装置74〜78が接続されている。ホスト70から送信されるリード/ライトコマンド要求に対して、ディスクアレイコントローラ73が、コマンド解釈、論理物理変換を行い、各ディスク装置74〜78に対して、データのリード/ライトアクセスを行う。

## 【0005】

【発明が解決しようとする課題】従来のディスクシステムでは、次の問題点がある。

(1) システム全体を高性能化するために高速アクセ

3

ス、大容量のディスク装置を揃えてディスクアレイを構成すると、高価アクセス、大容量のディスク装置は高価であるため、ディスク台数増加に伴ってディスクシステム全体のコストが上昇する。

(2) データ量の増加に対応させてディスク装置の台数を増加させると、ディスクアレイコントローラの負荷が増加し、システム全体のスループット性能の低下を招く。そこで、本発明の目的は、低コストで高スループット性能を実現するディスクシステムを提供することにある。

【0006】

【課題を解決するための手段】第1の観点では、本発明は、1台以上のホストコンピュータと、2台以上のディスクアレイとをファイバチャネルのループで接続したディスクシステムであって、シーケンシャルコマンドを優先的に処理するディスクアレイからランダムコマンドを優先的に処理するディスクアレイかを識別しうる識別IDを前記ディスクアレイの少なくとも2台に対して決定する識別ID決定手段と、前記ホストコンピュータから前記ディスクアレイに対して送信したコマンドがシーケンシャルコマンドかランダムコマンドかを区分するコマンド区分手段と、シーケンシャルコマンドにはシーケンシャルコマンドを優先的に処理するディスクアレイを割り当てると共にランダムコマンドにはランダムコマンドを優先的に処理するディスクアレイを割り当てるためにコマンドに対して前記識別IDを付加する識別ID付加手段とを具備したことを特徴とするディスクシステムを提供する。上記第1の観点のディスクシステムでは、ファイバチャネルのループに複数台のディスクアレイを接続し、各ディスクアレイに対してシーケンシャルコマンドを優先的に処理するかランダムコマンドを優先的に処理するかを機能分担させる。そして、ホストコンピュータからのコマンドがシーケンシャルコマンドかランダムコマンドかを区分し、その区分に応じたグループに該コマンドを処理させる。これにより、コマンドの特性に応じてディスクアレイのディスク装置を構成する（例えば、ランダムコマンドを優先的に処理するディスクアレイは基本的に低コストの低速アクセス・大容量のディスク装置で構成する）ことが可能となり、低コストで高スループット性能を実現することが出来る。さらに、データ量に応じてグループ分けを動的に変更でき、性能スケーラビリティを実現することが出来る。

【0007】第2の観点では、本発明は、1台以上のホストコンピュータと、1台以上のディスクアレイとをファイバチャネルのループで接続したディスクシステムであって、前記ディスクアレイの少なくとも1台は、2台以上のディスクアレイコントローラおよび2台以上のディスク装置を含み、前記ディスクアレイコントローラおよび前記ディスク装置をグループ化するディスクグループ構成手段と、シーケンシャルコマンドを優先的に処理

4

するグループからランダムコマンドを優先的に処理するグループかを識別しうる識別IDを前記グループに対して決定する識別ID決定手段と、前記ホストコンピュータから前記ディスクアレイに対して送信したコマンドがシーケンシャルコマンドかランダムコマンドかを区分するコマンド区分手段と、シーケンシャルコマンドにはシーケンシャルコマンドを優先的に処理するグループを割り当てると共にランダムコマンドにはランダムコマンドを優先的に処理するグループを割り当てるためにコマンドに対して前記識別IDを付加する識別ID付加手段とを具備したことを特徴とするディスクシステムを提供する。上記第2の観点のディスクシステムでは、ファイバチャネルのループに複数台のディスクアレイコントローラをもつディスクアレイを接続し、各ディスクアレイコントローラに対してディスク装置をグループ分けし、各グループに対してシーケンシャルコマンドを優先的に処理するかランダムコマンドを優先的に処理するかを機能分担させる。そして、ホストコンピュータからのコマンドがシーケンシャルコマンドかランダムコマンドかを区分し、その区分に応じたグループに該コマンドを処理させる。これにより、コマンドの特性に応じてグループのディスク装置を構成する（例えば、ランダムコマンドを優先的に処理するグループは基本的に高価な高速アクセスのディスク装置で構成するが、シーケンシャルコマンドを優先的に処理するディスクアレイは基本的に低コストの低速アクセス・大容量のディスク装置で構成する）ことが可能となり、低コストで高スループット性能を実現することが出来る。さらに、データ量に応じてグループ分けを動的に変更でき、性能スケーラビリティを実現することが出来る。

【0008】第3の観点では、本発明は、上記第1または第2の観点のディスクシステムにおいて、前記コマンド区分手段および前記識別ID付加手段を複数有し、前記コマンド区分処理および前記識別ID付加処理を多重実行可能としたことを特徴とするディスクシステムを提供する。上記第3の観点のディスクシステムでは、コマンド区分処理および識別ID付加処理を多重実行可能としたため、これらの処理がシステム性能に対してボトルネックになることを回避できる。したがって、ファイバチャネルのバンド幅を有効に利用した高スループットの運用が実現可能となる。

【0009】第4の観点では、本発明は、1台以上のホストコンピュータと1台以上の高速アクセスのディスクアレイと1台以上の高価なディスクアレイおよびI/OやMIOなどの低速アクセス・大容量のライブラリの少なくとも一つとをファイバチャネルのループで接続したディスクシステムであって、アクセス処理に従ってデータを前記高速アクセスのディスクアレイから前記高価なディスクアレイまたは前記低速アクセス・大容量のライブラリに移行するデータ移行手段と、移行後のデ

ータのアドレスを管理するデータベース手段とを具備したことを特徴とするディスクシステムを提供する。上記第4の観点のディスクシステムでは、記憶媒体の種類に注目し、アクセス頻度の違いによってデータを他の種類の記憶媒体に移動するため、各種類の記憶媒体が得意とするデータ提供方法でコマンド処理を行うことが出来る。したがって、低コストで高スループットの実現が実現可能となる。

【0010】第5の観点では、本発明は、上記第1から第4の観点のディスクシステムにおいて、複数本のファイバチャネルのループの利用率を検出し利用率的低いループを優先的に用いてコマンドを送信するループ選択手段をさらに具備したことを特徴とするディスクシステムを提供する。上記第5の観点のディスクシステムでは、ネットワークトラフィックの集中を回避することが出来る。

【0011】

【発明の実施の形態】以下、図を参照して本発明の実施の形態を説明する。なお、これにより本発明が限定されるものではない。

【0012】～第一実施形態～

第一実施形態は、ディスクアレイを複数化し、各ディスクアレイがシーケンシャル・ランダムに分類されたコマンドタイプに依ってコマンド処理を行う例である。図1は、本発明の第一実施形態にかかるディスクシステムの構成図である。このディスクシステム1は、ホストコンピュータ70と、ディスクアレイ86、91、96と、ネットワーク管理部97とを、ループ80、81で接続し、Storage Area Networkを構成したものである。前記ディスクアレイ86は、ディスクアレイコントローラA82と複数のディスク装置83～85とからなっている。前記ディスクアレイB91は、ディスクアレイコントローラB87と複数のディスク装置88～90とからなっている。前記ディスクシステムC96は、ディスクアレイコントローラC92と複数のディスク装置93～95とからなっている。

【0013】図2は、前記ネットワーク管理部97の構成図である。ネットワーク管理部97は、ループ80からコマンドを受信するポートA18、ループ81からコマンドを受信するポートB19、コマンドがシーケンシャル・ランダムに識別する処理を行うコマンド識別部10、識別対象のディスク装置の属性によってユーザがシーケンシャル・ランダムに分担処理を割り当てるためのディスクアレイコントローラ（82、87、92）を割り当てるためのコントローラ識別部11、前記コマンド識別部10で識別したコマンドを処理するディスクアレイコントローラ（82、87、92）を割り当てるためのコントローラ識別部12、そのコントローラ識別部12で発行したコントローラ識別部10

部11を各コマンドに付加するコントローラ識別部14付加部13、ホストコンピュータ70が持つ仮想アドレス（ループ80、81に接続された複数の記憶媒体を、1台の記憶媒体として見た場合のアドレス）を、ディスクアレイ（86、91、96）固有の論理アドレスに変換するアドレス変換部14、ループ80のトラフィックを検出してその利用率を算出するポートA利用率検出部16、ループ81のトラフィックを検出してその利用率を算出するポートB利用率検出部17、および、ポートA、Bの利用率を比較してコマンドをポートA18で送信するか又はポートB19で送信するかを決定するネットワークトラフィック比較部15とを具備して成る。

【0014】図3は、前記ネットワーク管理部97内のコマンド識別区分実行部10の詳細ブロック図である。コマンド識別区分実行部10は、ポートA18またはポートB19から受信したコマンドのアドレスをチェックするアドレス読み取り部30、1つのコマンドのアドレスを格納する前のコマンドのアドレス記憶部31、受信したコマンドのアドレスが連続か非連続か（等間隔か否か）を判定するアドレスの連続・非連続判定部32、および、連続したアドレスを持つコマンドにシーケンシャル・フラグを付加し非連続なアドレスを持つコマンドにランダム・フラグを付加するシーケンシャル・フラグまたはランダム・フラグ付加部33を具備して成る。

【0015】図4は、前記ネットワーク管理部97内のアドレス変換部14の詳細ブロック図である。アドレス変換部14は、ホストコンピュータ70がコマンドに割り当てる仮想アドレスを読み取る仮想アドレス読み取り/送信部34、その仮想アドレス読み取り/送信部34から受け取った仮想アドレスをディスクアレイ固有の論理アドレスに対応づける変換マップ35、および、その変換マップ35で変換された論理アドレスを受け取る論理アドレス受信部36を具備して成る。

【0016】図5および図6は、第一実施形態にかかるディスクシステム1の動作を示すフローチャートである。図5のステップ200では、ホストコンピュータ70がコマンドを発行する。ステップ201では、ホストコンピュータ70から発行されたコマンドがループ80またはループ81を流れる。ステップ202では、ループ80またはループ81を流れてきたコマンドを、ネットワーク管理部97のポートA18またはポートB19で受信する。ステップ203では、コマンド識別区分実行部10のアドレス読み取り部30で、受信したコマンドからアドレスを読み取り、アドレス連続・非連続判定部32で、前記コマンドのアドレス記憶部31から取り出したアドレスと前記読み取ったアドレスとを基に、コマンドのタイプがシーケンシャルかランダムかを識別し、シーケンシャル・フラグまたはランダム・フラグ付加部33で、シーケンシャルと識別したコマンドにシーケンシャル・フラグを付加し、ランダムと識別したコマンド

7

にランダム・フラグを付加する。ステップ204～206では、コントロール識別1D発行部12で、シーケンシャル・フラグの付加されたコマンドに対してシーケンシャル・コマンドを処理するディスクアレイコントローラの識別1Dを生成し(205)、ランダム・フラグの付加されたコマンドに対してランダム・コマンドを処理するディスクアレイコントローラの識別1Dを生成する(206)。ステップ207では、コントロール識別1D付加部13で、コマンドにコントロール識別1Dを付加する。なお、ステップ208に示すように、ディスクシステム構成時に、ユーザが、アクセス速度や容量等のディスク装置の属性を考慮して、どのディスクアレイコントローラ(82, 87, 92)をシーケンシャル・コマンド処理用とするか及びランダム・コマンド処理用とするか及び何番のコントロール識別1Dを持させるかを、ネットワーク管理部97において、決定しておく。

【0017】図6のステップ300では、アドレス変換部14の仮想アドレス読み取り 送信部34で、コマンドの仮想アドレスを読み取り、変換部35で、各コントロール識別1Dに対応したディスクアレイ(86, 91, 96)の論理アドレスに変換する。ステップ301～307では、ネットワーク管理部97のポートA利用率検出部16およびポートB利用率検出部17でループ80及びループ81の利用率を検出し(302, 303)、ネットワークラフィック比較部15でループ80, 81の利用率を比較し(301)、利用率がポートA<ポートBまたはポートA=ポートBならポートA1からコマンドを送信し(305)、利用率がポートA>ポートBならポートB19からコマンドを送信する(307)。ステップ308では、ループ80または81を介して送信されたコマンドを、該コマンドが持つコントロール識別1Dに対応するディスクアレイコントローラ(82, 87, 92)が受信する。ステップ309では、コマンドを受信したディスクアレイコントローラ(82, 87, 92)が、論理アドレスをディスク装置の物理アドレスに変換する等のコマンドの解釈を行う。ステップ310では、コマンドがリードならステップ311へ進み、ライトならステップ315へ進む。ステップ311では、キャッシュヒットの有無を判定し、キャッシュヒットならステップ312へ進み、キャッシュミスならステップ313へ進む。ステップ312では、要求データをキャッシュからリードする。そして、ステップ314へ進む。ステップ313では、要求データをディスク装置からリードする。ステップ314では、リードしたデータを、ループ80またはループ81を介して送信し、ホストコンピュータ7が受信する。ステップ315では、データをディスク装置にライトする。

【0018】例えば、ユーザがコントロール識別1Dユーザ決定部11においてディスクアレイコントローラA82をシーケンシャルコマンド処理用とし、ディスクア

8

レイコントローラB87をランダムコマンド処理用とし、ディスクアレイコントローラC92をシーケンシャルコマンド処理用(ディスクアレイA86の容量を超えてライトコマンドを受信するような場合に用いる)とし、コントロール識別1Dを決定した場合、ディスクアレイA86のディスク装置83～85およびディスクアレイC96のディスク装置93～95に対してはシーケンシャル処理のみを行うため、アクセス性能に依存しないディスク装置の利用が可能となる。一方、ディスクアレイB91のディスク装置88～90に対してはランダム処理のみを行うため、アクセス性能とキャッシュ性能に特化したディスク装置の利用が可能となる。

【0019】上記第一実施形態のディスクシステム1によれば、より高性能なディスク装置はランダム専用用を用い、平凡な安価なディスク装置はシーケンシャル用を用いることで、高性能かつ低コストなディスクシステムを実現できる。

#### 【0020】第二実施形態一

第二実施形態は、1台のディスクアレイに複数のディスクアレイコントローラを増設する例である。図7は、本発明の第二実施形態にかかるディスクシステムの構成図である。このディスクシステム2は、ホストコンピュータ70と、コントロール増設ディスクアレイ98と、ネットワーク管理部99とを、ループ80, 81で接続し、Storage Area Networkを構成したものである。前記コントロール増設ディスクアレイ98は、複数のディスクアレイコントローラ82, 87, 92と複数のディスク装置83～95とからなっている。複数のディスク装置83～95は、複数のディスクアレイコントローラ82, 87, 92に対応したディスクグループA, B, Cにグループ分けされている。

【0021】図8は、前記ネットワーク管理部99のブロック図である。ネットワーク管理部99は、ループからコマンドを送受信するポートA18、ポートB19、コマンドがシーケンシャルかランダムかを識別するコマンド識別区分実行部10、識別したシーケンシャル・コマンドを一時保持するシーケンシャル用コマンドキュー42、識別したランダム・コマンドを一時保持するランダム用コマンドキュー43、前記コマンドキュー42, 43に溜まったコマンドの数を比較してディスク装置83～95をシーケンシャルコマンド用グループ、ランダムコマンド用グループにグループ分けするディスクグループ構成部41、ディスク装置の属性によってシーケンシャル・ランダムとの分担処理を割り当てるためのディスクアレイコントローラ(82, 87, 92)のコントロール識別1Dをユーザが決定するコントロール識別1Dユーザ決定部11、識別したコマンドを処理するディスクアレイコントローラ(82, 87, 92)を割り当てるためのコントロール識別1Dを発行するコントロール識別1D発行部45、そのコントロール識別1D発行部

9

45で発行したコントローラ識別IDを各コマンドに付加するコントローラ識別ID付加部13、前記ディスクグループを考慮してホストコンピュータ70が持つ仮想アドレスをディスクアレイ固有の論理アドレスに変換するアドレス変換部44、ループ80のトラフィックを検知してその利用率を算するポートA利用率検出部16、ループ81のトラフィックを検知してその利用率を算するポートB利用率検出部17、検出したループ80、81の利用率の結果を比較してコマンドをポートA18で送信するかポートB19で送信するかを決定するネットワークトラフィック比較部15、および、ループ初期化処理でループ80、81に接続されたデバイスを認証するためのループ初期化に伴うデバイスチェック部40を具備して成る。

【0022】図9は、前記ループ初期化に伴うデバイスチェック部40の詳細図である。ループ初期化に伴うデバイスチェック部40は、ポートA18、ポートB19から初期化シーケンスのフレームを読み取るLIP (Loop Initialization Procedure) における初期化シーケンスのフレーム読み取り部50、読み取ったフレームから新旧デバイスのポートアドレスを比較するポートアドレス比較部51、すでにループを構成しているデバイスのポートアドレスを保持する既存ポートアドレス格納部52、および、新たにループに接続したデバイスのポートアドレスを保持する新規ポートアドレス格納部53を具備して成る。

【0023】図10は、前記ディスクグループ構成部41の詳細図である。ディスクグループ構成部41は、前記シーケンシャル用コマンドキュー42に溜まったコマンドの数をカウントするシーケンシャル用キューカウンタ54、前記ランダム用コマンドキュー43に溜まったコマンドの数をカウントするランダム用キューカウンタ55、前記キューカウンタ54、55のカウント値を比較するカウンタ比較部56、比較結果からディスク装置83~95のどれをシーケンシャルグループに割り当てるか決定するシーケンシャルグループ決定部57、および、どれをランダムグループに割り当てるか決定するランダムグループ決定部58を具備して成る。

【0024】図11~図13は、第2実施形態にかかるディスクシステム2の動作を示すフローチャートである。ステップ400では、ネットワーク管理部99のループ初期化に伴うデバイスチェック部40で、初期化シーケンスのフレームを読み取る。ステップ401では、コントローラ増設による新規ポートアドレスが検出された場合はステップ402へ進み、そうでない場合はステップ402をスキップする。ステップ402では、新規ポートアドレスをコントローラ識別ID発行部45に通知する。以上の前置き処理の後、ホストコンピュータ70がコマンドを発行するのを待つ。

【0025】ステップ403では、ホストコンピュータ

10

70がコマンドを発行する。ステップ404では、発行されたコマンドがループ80またはループ81を流れる、ステップ405では、流れてきたコマンドをネットワーク管理部99のポートA18またはポートB19で受信する。ステップ406では、コマンド識別区分実行部10で、受信したコマンドからコマンドのアドレスの連続・非連続性を判定し、コマンドのタイプを識別し、シーケンシャルと識別したコマンドにシーケンシャル・フラグを付加し、ランダムと識別したコマンドにランダム・フラグを付加する。ステップ407では、シーケンシャル・フラグを付加したならステップ408へ進み、ランダム・フラグを付加したならステップ412へ進む。ステップ408では、シーケンシャル用コマンドキュー42にコマンドを格納する。ステップ409では、コントローラ識別ID発行部45で、シーケンシャル用コマンドキュー42の先頭のコマンドに対して、シーケンシャル・コマンドを処理するディスクアレイコントローラ(82, 87, 92)のコントローラ識別IDを生成する。ステップ410では、コントローラ識別ID付加部13で、生成されたコントローラ識別IDをコマンドに付加する。そして、図12のステップ502へ進む。

【0026】ステップ412では、ランダム用コマンドキュー43にコマンドを格納する。ステップ413では、コントローラ識別ID発行部45でランダム用コマンドキュー43の先頭のコマンドに対して、ランダム・コマンドを処理するディスクアレイコントローラのコントローラ識別IDを生成する。そして、前記ステップ410へ進む。なお、ステップ415に示すように、ディスクシステム構成時に、ユーザが、アクセス速度や容量等のディスク装置の属性を考慮して、どのディスクアレイコントローラ(82, 87, 92)をシーケンシャル・コマンド処理用とするか及びランダム・コマンド処理用とするか及び何番のコントローラ識別IDを付与するかを、ネットワーク管理部97において、決定しておく。

【0027】図12のステップ500、501は、前記ステップ407、412でコマンドをキュー42、43に格納した後、実行される。ステップ500では、ディスクグループ構成部41のシーケンシャル用キューカウンタ54とランダム用キューカウンタ55の値をカウンタ比較部56で比較する。ステップ501では、シーケンシャルグループ決定部57でシーケンシャルグループに用いるディスク装置をグループ化し、ランダムグループ決定部58でランダムグループに用いるディスク装置をグループ化する。以上のステップ500、501で、ステップ502が実行される準備ができる。

【0028】ステップ502では、コマンドの仮想アドレスをアドレス変換部44で読み取り、ディスク装置のグループ化を反映して、各コントローラ識別IDに対応

したディスクアレイの論理アドレスに変換する。ステップ503～509では、ネットワーク管理部99のポートA利用率検出部106およびポートB利用率検出部17でループ80及びループ81の利用率を検出し(504、505)、ネットワークラフィック比較部15でループ80、81の利用率を比較し(503)、利用率がポートA<ポートBまたはポートA=ポートBならポートA18からコマンドを送信し(507)、利用率がポートA>ポートBならポートB19からコマンドを送信する(509)。そして、図13のステップ600へ進む。

【0029】図13のステップ600では、ループ80またはループ81を介して送信されたコマンドを、該コマンドが持つコントロール識別IDに対応するディスクアレイコントローラ(82、87、92)が受信する。ステップ601では、コマンドを受信したディスクアレイコントローラ(82、87、92)が、論理アドレスをディスク装置の物理アドレスに変換する等のコマンドの解釈を行う。ステップ602では、コマンドがリードならステップ603へ進み、ライトならステップ607へ進む。ステップ603では、キャッシュヒットの有無を判定し、キャッシュヒットならステップ604へ進み、キャッシュミスならステップ605へ進む。ステップ604では、要求データをキャッシュからリードする。そして、ステップ606へ進む。ステップ605では、要求データをディスク装置からリードする。ステップ606では、リードしたデータを、ループ80またはループ81を介して送信し、ホストコンピュータ70が受信する。ステップ607では、データをディスク装置にライトする。

【0030】上記第二実施形態のディスクシステム2によれば、より高性能なディスク装置はランダム専用を用い、平凡な安価なディスク装置はシーケンシャル専用を用いることで、高性能かつ低コストなディスクシステムを実現できる。

### 【0031】—第三実施形態—

第三実施形態は、複数のネットワーク管理部を持つ例である。図14は、本発明の第三実施形態にかかるディスクシステムの構成図である。このディスクシステム3は、ホストコンピュータ70と、ディスクアレイ86、91、96と、ネットワーク管理部97、100、101とを、ループ80、81で接続し、Storage Area Networkを構成したものである。前記ディスクアレイ86は、ディスクアレイコントローラ82と複数のディスク装置83～85とからなっている。前記ディスクアレイB91は、ディスクアレイコントローラB87と複数のディスク装置B88～90とからなっている。前記ディスクシステムC96は、ディスクアレイコントローラC92と複数のディスク装置93～95とからなっている。

【0032】上記第三実施形態のディスクシステム3によれば、複数のネットワーク管理部97、100、101で処理を分担することで、シーケンシャル/ランダムのコマンド区分に要する処理を同一時間内に多重実行可能となり、コマンドの区分等の処理がシステム性能に対してボトルネックになることを回避できる。したがって、ファイバチャネルのバンド幅を有効に利用した高スループットの運用が実現可能となる。

### 【0033】—第四実施形態—

10 第四実施形態は、異なる特性の記憶媒体にデータを移行するディスクシステムの例である。図15は、本発明の第四実施形態のディスクシステムの構成図である。このディスクシステム4は、ホストコンピュータ70と、高速アクセスのRAID1(ミラー)の構成を持つディスクアレイ104と、高信頼性だがバリエーションのためにややアクセス速度が落ちるRAID5(パリティストライピング)の構成を持つディスクアレイ105と、大容量であるが低速アクセスのDVDやMOで実現されるライブラリ106と、ネットワーク管理部102とを、ループ80、81で接続し、Storage Area Networkを構成したものである。前記ネットワーク管理部102は、ある記憶媒体のデータを他の記憶媒体に移行するためのマッピングデータを保持するデータベース103を有している。

【0034】図16は、前記データベース103の例示図である。このデータベース103は、データ名と、移動先アドレスと、元アドレスと、データサイズと、移動先とを管理している。

30 【0035】図17は、第四実施形態にかかるディスクシステム4の動作を示すフローチャートである。ステップ700では、ホストコンピュータ70が、RAID1のディスクアレイ104にコマンドを発行する。ステップ701では、RAID1のディスクアレイ104でコマンドの処理を行う。ステップ702では、RAID1のディスクアレイ104に対するアクセス頻度が低い場合は前記ステップ700に戻る(RAID1のディスクアレイ104で引き続き処理を行う)が、アクセス頻度が高くなったら、ステップ703へ進む。

40 【0036】ステップ703では、RAID1のディスクアレイ104のデータの中で特に参照する回数が多いデータを、RAID5のディスクアレイ105に移動する。このとき、データの移動毎に、移動したデータのアドレスをデータベース103に記録して、以降のコマンドアクセスで参照できるようにする。ステップ704では、RAID1のディスクアレイ104及びRAID5のディスクアレイ105でコマンドの処理を行う。ステップ705では、RAID5のディスクアレイ105に対するアクセス頻度が低い場合は前記ステップ708に進み、アクセス頻度が高くなったら、ステップ706へ進む。ステップ708では、ホストコンピュータ70



が、RAID1のディスクアレイ104及びRAID5のディスクアレイ105にコマンドを発行する。そして、RAID1のディスクアレイ104及びRAID5のディスクアレイ105で引き続き運用を行う。

【0037】ステップ706では、RAID5のディスクアレイ105のデータの内特に参照する回数の少ないデータをライブラリ106に移動する。このとき、データの移動毎に、移動したデータのアドレスをデータベース103に記録して、以降のコマンドアクセスで参照できるようにする。ステップ707では、RAID1のディスクアレイ104及びRAID5のディスクアレイ105及びライブラリ106でコマンドの処理を行う。ステップ709では、ホストコンピュータ70が、RAID1のディスクアレイ104及びRAID5のディスクアレイ105及びライブラリ106にコマンドを発行する。そして、RAID1のディスクアレイ104及びRAID5のディスクアレイ105及びライブラリ106で引き続き運用を行う。

【0038】上記第四実施形態のディスクシステム4によれば、Storage Area Networkを構成する記憶媒体の種類に注目し、アクセス頻度の違いによってデータを他の種類の記憶媒体に移動するため、各種類の記憶媒体が得意とするデータ提供方法でコマンド処理を行うことが出来る。

【0039】

【発明の効果】本発明のディスクシステムによれば、次の効果が得られる。

(1) リード/ライトコマンドをシーケンシャルランダムで区分してターゲットのディスク装置に送信することで、高価格で高速アクセス、低価格で大容量といったディスク装置が固有に持つ属性に合った運用が可能になり、低コストで高スループット性能を実現することが出来る。また、複数のディスクアレイコントローラにより負荷分散するため、ディスク装置の台数が多い大規模なディスクシステムにおいても、コントローラネックのない性能スケーラビリティを示す。

(2) 複数のディスクアレイコントローラがそれぞれ管理するディスク装置の構成を動的に変更できる。

(3) コマンド区分処理および識別ID付加処理がシステム性能に対してボトルネックになることを回避できる。

(4) データを種類の異なる記憶媒体に移動し、各種類の記憶媒体が得意とするデータ提供方法でコマンド処理を行うことが出来る。

(5) ネットワークトラフィックの集中を回避することが出来る。

【図面の簡単な説明】

【図1】本発明の第一実施形態のディスクシステムの構成図である。

【図2】図1に示されたネットワーク管理部の構成図で

ある。

【図3】図2に示されたコマンド識別区分実行部の構成図である。

【図4】図2に示されたアドレス変換部の構成図である。

【図5】本発明の第一実施形態の動作を示すフローチャートである。

【図6】図5の続きのフローチャートである。

【図7】本発明の第二実施形態のディスクシステムの構成図である。

【図8】図7に示されたネットワーク管理部の構成図である。

【図9】図8に示されたループ初期化に伴うデバイスチェック部の構成図である。

【図10】図8に示されたディスクグループ構成部の構成図である。

【図11】本発明の第二実施形態の動作を示すフローチャートである。

【図12】図11の続きのフローチャートである。

【図13】図12の続きのフローチャートである。

【図14】本発明の第三実施形態のディスクシステムの構成図である。

【図15】本発明の第四実施形態のディスクシステムの構成図である。

【図16】図15に示されたデータベースの構成図である。

【図17】本発明の第四実施形態の動作を示すフローチャートである。

【図18】従来のディスクシステムの一例の構成図である。

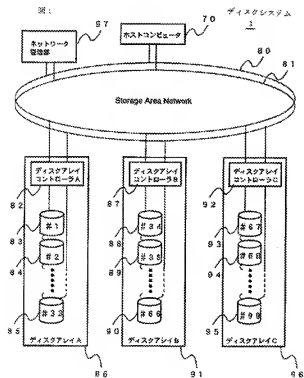
【符号の説明】

- 1、2、3、4・・・ディスクシステム、
- 10・・・コマンド識別区分実行部、
- 11・・・コントローラ識別IDユース決定部、
- 12・・・コントローラ識別ID発行部、
- 13・・・コントローラ識別ID付加部、
- 14、44・・・アドレス変換部、
- 15・・・ネットワークトラフィック比較部、
- 16・・・ポートA利用率検出部、
- 17・・・ポートB利用率検出部、
- 18・・・ポートA、
- 19・・・ポートB、
- 30・・・アドレス読み取り部、
- 31・・・前のコマンドのアドレス記憶部、
- 32・・・アドレス連結/非連続判定部、
- 33・・・シーケンシャル・フラグまたはランダム・フラグ付加部、
- 34・・・仮想アドレス読み取り・送信部、
- 35・・・変換マップ、
- 36・・・論理アドレス受領部、

15

- 40・・・ループ初期化に伴うデバイスチェック部。  
 41・・・ディスクグループ構成部。  
 42・・・シーケンシャル周コマンドキュー。  
 43・・・ランダム周コマンドキュー。  
 50・・・LIPにおける初期化シーケンスのフレーム読み取り部。  
 51・・・ポートアドレス比較部。  
 52・・・既存ポートアドレス格納部。  
 53・・・新規ポートアドレス格納部。  
 54・・・シーケンシャル用キューカウンタ。  
 55・・・ランダム用キューカウンタ。  
 56・・・カウンタ比較部。  
 57・・・シーケンシャルグループ決定部。  
 58・・・ランダムグループ決定部。

【図1】



【図16】

図16

データベース

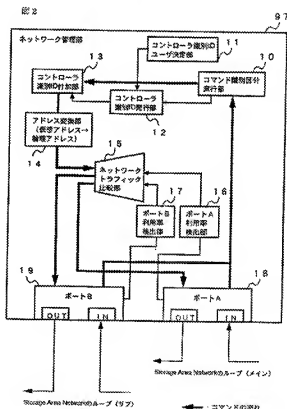
182

ファイル名	移動先アドレス	元アドレス	サイズ	移動先
AAAAA	YYYYY	XXXXX	ZZZ	8456
BBBBB	PPPPP	QQQQQ	WWWWW	547 23

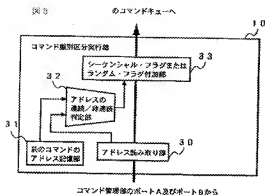
16

- 70・・・ホストコンピュータ。  
 80、81・・・ループ。  
 82、87、92・・・ディスクアレイコントローラ。  
 83～85、88～90、93～95・・・ディスク装置。  
 86、91、96・・・ディスクアレイ。  
 97、99、100、101、102・・・ネットワーク管理部。  
 98・・・コントローラ増設ディスクアレイ。  
 103・・・データベース。  
 104・・・RAID1ディスクアレイ。  
 105・・・RAID5ディスクアレイ。  
 106・・・DVD、MOライブラリ

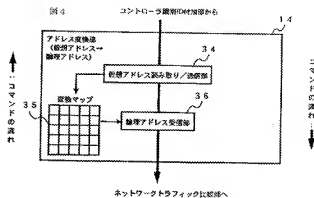
【図2】



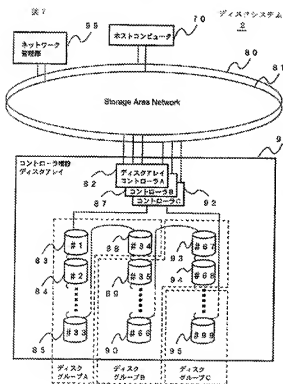
【図3】



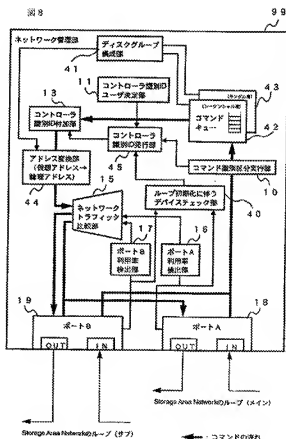
【図4】



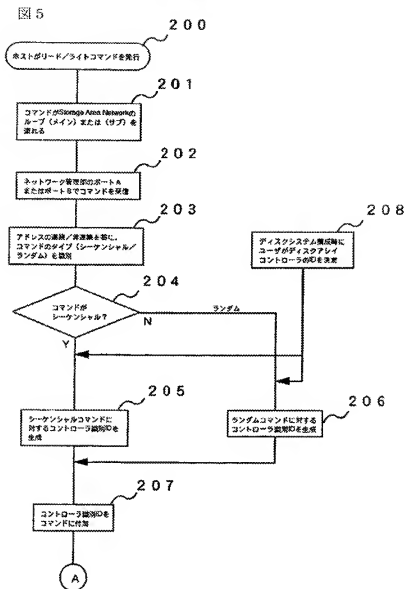
【図7】



【図8】

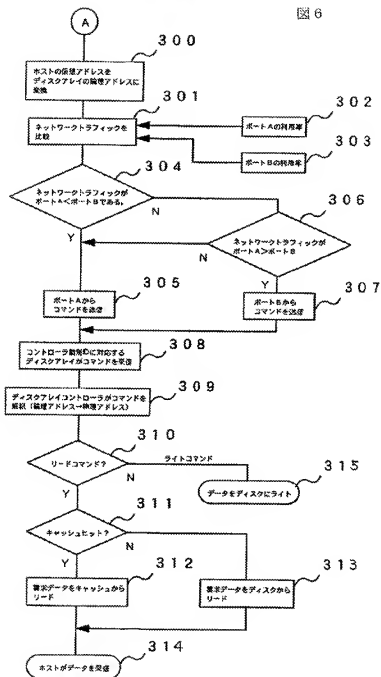


【図5】

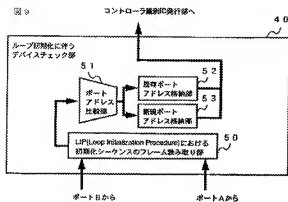


【図6】

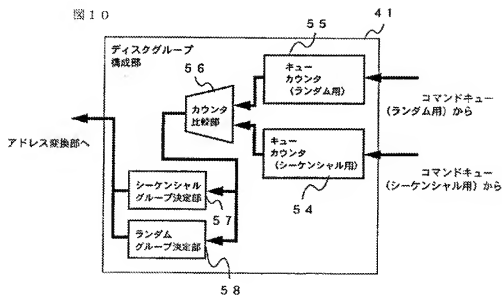
図6



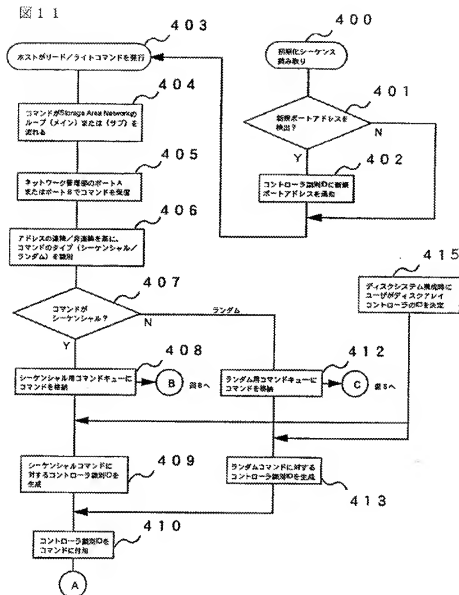
【図9】



【図10】

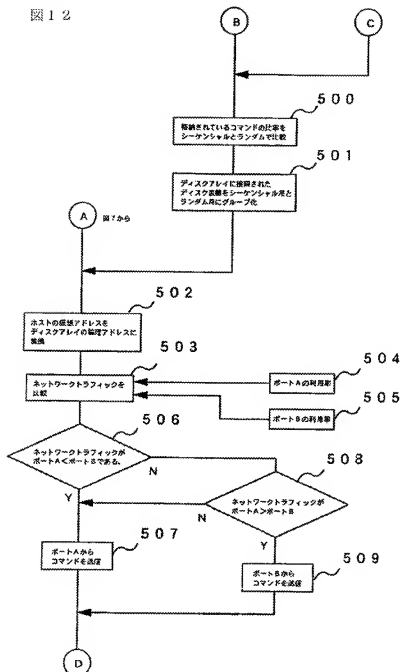


【図11】



【図12】

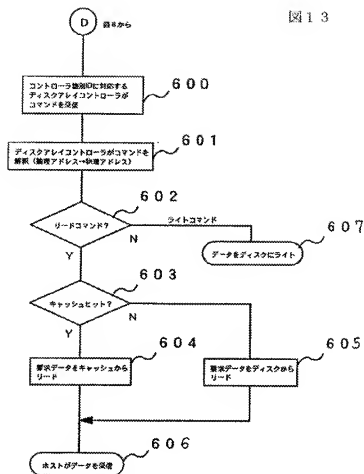
図12



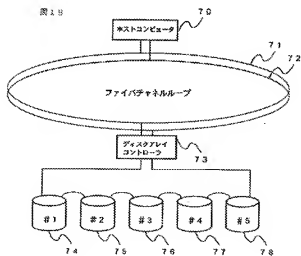


【図13】

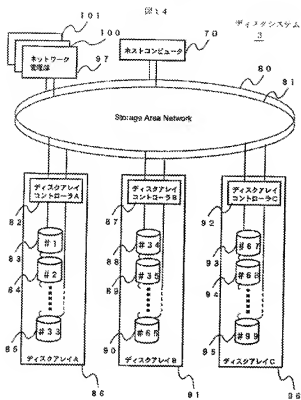
図13



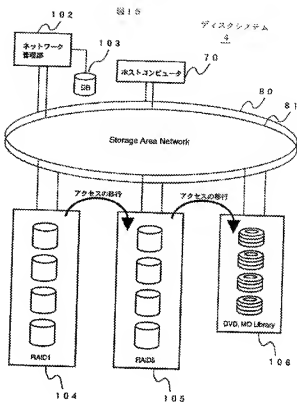
【図18】



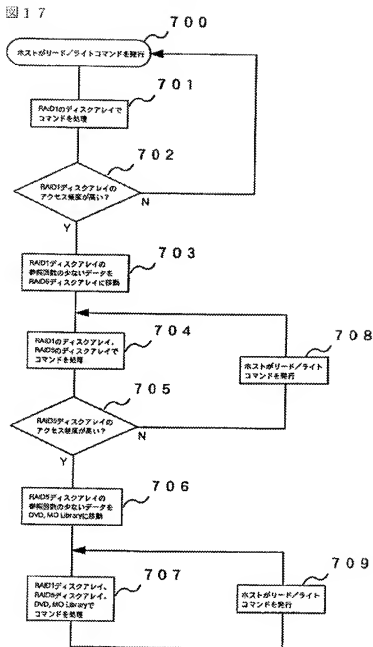
【図14】



【図15】



【図17】



フロントページの続き

(72) 発明者 八木沢 晋哉  
 神奈川県川崎市麻生区王禅寺1099番地 様  
 株式会社日立製作所システム開発研究所内  
 (72) 発明者 山本 政行  
 神奈川県川崎市麻生区王禅寺1099番地 様  
 株式会社日立製作所システム開発研究所内

(72) 発明者 ▼高▲本 賢一  
 神奈川県小田原市国府津2880番地 株式会社  
 日立製作所ストレージシステム事業部内  
 Fターム(参考) 5B065 BA01 BA03 BA05 CA15 CA17  
 CA30